DOI: https://doi.org/10.52756/ijerr.2025.v47.007 ***Int. J. Exp. Res. Rev.*, Vol. 47: 84-92 (2025)**

Original Article | *Peer Reviewed* | *Open Access*

International Journal of Experimental Research and Review (IJERR)
© Copyright by International Academic Publishing House (IAPH)
ISSN: 2455-4855 (Online)
www.iaph.in
ISSN 2455-4855

A Comparative Study of Accident Black Spot Identification: A Crucial Analysis Using Machine Learning and Statistical Approach

Amrita Sarkar[1]*and Vandana Bhattacharjee[2]

[1]Department of CSE, Birla Institute of Technology, Mesra (Lalpur), Ranchi-834001, Jharkhand, India; [2]Department of CSE, Birla Institute of Technology, Mesra, Ranchi-835215, Jharkhand, India

E-mail/Orcid Id:

AS, amritasarkar@bitmesra.ac.in, https://orcid.org/0000-0001-7415-2822; *VB,* vbhattacharya@bitmesra.ac.in, https://orcid.org/0000-0002-0680-2691

Article History:

Received: 16th Sep., 2024
Accepted: 24th Dec., 2024
Published: 30th Apr., 2025

Keywords:

Black Spot Identification, Urban Traffic Accident Prediction, Traffic Safety Management, Statistical Methods, Machine Learning Approaches.

How to cite this Article:

Amrita Sarkar and Vandana Bhattacharjee (2025). A Comparative Study of Accident Black Spot Identification: A Crucial Analysis Using Machine Learning and Statistical Approach. *International Journal of Experimental Research and Review*, *47*, 84-92.
DOI: https://doi.org/10.52756/ ijerr.2025.v47.007

Abstract: Identifying black spots in traffic accidents is critical for the urban traffic accident prediction model and better traffic safety management. Several strategies have been used to identify accidents on urban roadways. Some employ Statistical methodologies, while others, more recently, employ Machine Learning approaches. Fuzzy Logic-based algorithms are ideal for predicting car collisions because they can quickly generate complex non-linear relationships between data. A fully fuzzy-based technique, on the other hand, would provide the required user-defined knowledge. Neural Networks are suitable for this purpose because of their sophisticated learning capabilities. As a result, when the driving attributes are employed as a predictor variable in the accident prediction model, a neuro-fuzzy technique will produce the intended outcome.

Introduction

Accident black spots, also known as high-risk locations or critical road segments, are areas on road networks where traffic accidents are significantly higher than other road parts. Identifying and mitigating these black spots is crucial for improving road safety and reducing the number of traffic-related injuries and fatalities. Accurate identification of accident black spots involves analyzing historical crash data, traffic flow patterns, road geometry, environmental factors, and other relevant variables. Transportation authorities can implement targeted interventions to enhance safety measures and prevent future incidents by understanding the factors contributing to accidents at specific locations. Accident black spot identification typically begins with collecting and analyzing crash data from various sources, such as police reports, hospital records, and insurance claims. This data is then spatially mapped to identify clusters or concentrations of accidents along roadways. Several statistical and data mining techniques are commonly used to analyze crash data and identify black spots. These include spatial analysis, cluster analysis, regression modelling, and machine learning algorithms. These methods help to uncover underlying patterns and trends in crash data, allowing transportation planners and engineers to prioritize resources and interventions effectively. Once identified, accident black spots can be addressed through a range of countermeasures and safety improvements. These may include traffic signal optimization, signage and pavement markings, road geometry modifications, speed limit adjustments, enforcement strategies, and public awareness campaigns. Efforts to identify and mitigate accident black spots are essential components of comprehensive road safety management strategies at the local, regional, and national levels. By targeting resources and interventions to areas with the highest risk of accidents, transportation agencies can make significant strides toward reducing injuries and fatalities on roadways and improving overall traffic safety for all road users. Extensive work has been done on the identification of Black Spots. There is hardly any universal definition of a Black Spot. Black Spots were identified from the sites after a detailed analysis of the risk estimated and the chances of an accident occurring at a particular point. To overcome this, the safety of the Black Spot may be improved by adopting road safety improvements and altering the geometrical and

***Corresponding Author:** amritasarkar@bitmesra.ac.in

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

environmental characteristics of the identified location within a road network.

Literature Review

Statistical models

Researchers have tried using various methods to identify traffic hot spots. The easiest method to locate Black Spots is to grade and classify them according to accident frequency and accident data in descending order. This approach has particular benefits; if the traffic hotspots are appropriately identified, there are several opportunities for growth and development (Norden et al., 1956; Brown et al., 1992; Heydecker and Wu, 2001; Kim et al., 2006; Wan et al., 2021; Karamanlis et al., 2023). For attribution, this conventional method necessitates the collection of historical facts and records. The research focuses on overcoming this scarcity and inventing and optimizing identification systems (Tamburri and Smith, 1970; Petridou and Moustak, 2000; Korhonen, 2018; Rashid and Kausik, 2024; Alhosani and Alhashmi, 2024).

Historically, researchers have tested various statistical models to calculate the rates of accidents and their frequencies at a particular site within an interval. The following are a few models that find application, with a thorough discussion available in Nassar (1996). Simple models based on mean and variance have been employed by Foldvary (1979) and Jovanis and Delleur (1983). These types studied the variations in accident rates under different circumstances but could not incorporate the effect of risk factors.

Multiple Linear Regression is defined as finding the position of an unknown location together with a network utilizing references from well-known places (Vonderohe and Chou, 1997). It can also technically assist in determining an unknown location on a transportation network, any distance along that way from a predetermined datum point, and optionally, an offset from the path can also be provided. Multiple Linear Regression models were employed for the same purpose by Oppe (1979) and Ceder and Livneh (1982). In these proposals, the number of accidents was made the dependent variable, and speed and traffic volume as the independent variables. Accident occurrence was assumed to be distributed in these models generally. These models' flaw is that they require the distributional property to properly account for the random and discrete car collision occurrences on the road (Geurts and Wets, 2003).

The inputs for Black Spot detection should be relevant, which the Poisson distribution approach can quickly determine. To test the Poisson distribution approach on traffic accidents, recent research looked at areas with many Accident Black Spots (Kristen and Mark, 2002; Meel and Safr, 2008). Poisson distribution describes four categories of dangerous road stretch susceptible to accidents (Huang et al., 2008), accident rates, and futuristic modelling (Caliendo et al., 2007; Mitra and Washington, 2007). Poisson assumptions can be used to arrange crash densities (Li et al., 2008) and time series (Quddus, 2008; Cheng et al., 2013).

The Poisson distribution was created to handle occurrences that happened randomly and to test repeatedly for events with a low likelihood of happening. Through analysis of Black Spot occurrence, it was observed that Black Spots formed by assembling independently occurring accidents in a cluster. This observation shows that Poisson distribution in such cases concerning randomly occurring traffic accidents is a highly relevant statistical tool for determining Black Spots.

The Poisson-gamma generalized linear model is presently accepted as one of the best models in this field. The Poisson mean can fluctuate between sites beyond what can be explained by variations in geographical parameters. According to literature reviews, most researchers explored individual accident locations or vehicle types to model accident rates, while relatively few have explored general accident involvement models.

A Poisson log-linear model was used by Saccomanno and Buyco (1988) and Blower et al. (1993) to provide vital insights into fluctuations in accident rates. The Poisson regression model is the best choice for information processing with a high percentage of zero counts. Because it ignores extra-Poisson volatility in observed accident counts, this model is inappropriate for measuring the number of traffic accidents.

Several authors have used negative binomial regression models to address the extra-Poisson variation problem (Maycock and Hall, 1984; Hauer and Persaud, 1987; Miaou, 1994; Shankar et al., 1995; Maher & Summersgill, 1996; Kulmala, 1995; Hauer, 1997; Tunaru, 1999; Abdel-Aty and Radwan, 2000). Miaou (1994) generated two negative binomial solutions, one by the maximum likelihood technique and the other by the moments' method. This maximum likelihood model performed better than the Poisson regression model while forecasting crashes with over-dispersion. Persaud et al. (1999) further advanced the negative binomial model, incorporating an empirical Bayesian adjustment to improve the predictability that he argued was low historically.

Hauer and Persaud (1987) further focused on extra-Poisson variation in road accident data. They worked on Maycock and Hall's (1984) generalized linear model, which associated the mean of the Poisson distribution with several location variables to account for organized changes in crash frequency between identified spots. The Poisson mean might differ between places, as suggested in their model by (Geurts and Wets, 2003). Maher and Summersgill (1996) supported Hauer and Persaud (1987). Poisson-gamma generalized linear model. Kulmala (1995) and Hauer (1997) broadly accepted this paradigm. This model has since been improved. By modelling the indicators in the gamma distribution and the regression parameters as random factors, gave the model a new dimension. Without explicitly modelling for dispersion effects, several techniques employed a simplistic linear model with a negative binomial error structure (Abdel-Aty and Radwan, 2000).

DOI: https://doi.org/10.52756/ijerr.2025.v47.007

Table 1. Comparison table for Statistical model.

Statistical Model	Description	Application	References
Grading and Classification	Identifies Black Spots based on accident frequency and data, facilitating growth and development opportunities.	Identifying high-risk areas by analyzing historical accident data.	Brown et al. (1992); Heydecker and Wu (2001); Kim et al. (2006); Norden et al. (1956)
Multiple Linear Regression	Predicts accident occurrences using variables like speed and traffic volume.	Determining accident likelihood based on traffic flow and environmental factors.	Oppe (1979); Ceder and Livneh (1982); Vonderohe and Chou (1997)
Poisson Distribution	Models accident occurrence based on the assumption of random events with a constant rate.	Predicting accident rates and arranging crash densities.	Kristen and Mark (2002); Meel and Safr (2008); Huang et al. (2008); Li et al. (2008); Quddus (2008)
Poisson-Gamma Generalized Linear Model	Incorporates variability in accident rates across different locations.	Provides a comprehensive model for understanding accident rates.	Maycock and Hall (1984); Hauer and Persaud (1987); Maher and Summersgill (1996); Kulmala (1995); Abdel-Aty & Radwan (2000)
Negative Binomial Regression	Addresses over-dispersion in accident data, providing more accurate predictions.	Modelling accident occurrences with variability beyond the mean.	Miaou (1994); Persaud et al. (1999); Hauer and Persaud (1987); Vistisen (2002); Hauer (2001)
Empirical Bayesian Approaches	Evaluates safety at specific locations by combining crash history and projected accidents on comparable sites.	Identifying high-risk sites and suggesting safety enhancements.	Hauer et al. (1988); Higle and Witkowski (1988)
Logistic Regression	Predicts accident severity based on categorical variables.	Models accident severity to understand potential outcomes.	Saccomanno and Buyco (1988); Blower et al. (1993)
Survival Theory Models	Predicts the likelihood of vehicle collisions at specific times, assuming survival until those times.	Requires specific data for predicting accident probabilities over time.	Jovanis and Chang (1989)

In the preceding model descriptions, reported numbers of incidents throughout the observation period were only considered when location attributes were assumed to be constant during this period. However, many properties, particularly traffic flow, have been shown to alter over time. It might be broken into sub-periods to provide a more complete explanation of the observation period. However, given more research, site-specific factors should be represented in location characteristics in different sub-periods at the exact location because they are not independent (Maher and Summersgill, 1996). Since accident numbers are not independently negative binomially distributed, the above technique raises the issue of estimating the model, for example, keeping separate records for each year's accident count.

Persaud (1990) assumed simultaneous independence of yearly accident counts, which proved impossible. Vistisen (2002) used disaggregated accident models for one-year sub-periods if annual accident counts depend on the same place. Hauer (2001) used a hierarchical Poisson-gamma distribution to characterize the random fluctuations in accident frequencies with the fixed and dispersion portions of the Poisson mean.

Empirical Bayesian approaches (EB) have also been employed to identify Black Spots. When both crash history and projected accidents on comparable sites are evaluated, EB methods have been judged to satisfy the criterion of a more remarkable ability to detect "high-risk" sites. Empirical Bayes (EB) approaches were used by Hauer et al. (1988) to evaluate the safety at signalized junctions. Persaud (1990) also assessed the collision hazard of Ontario Road sections and identified potential safety enhancements. Higle and Witkowski (1988) have proposed a different EB approach based on crash rates. Most of these researchers achieved favourable results in identifying hotspots; however, the range of conditions within studies could have been much higher.

Jovanis and Chang (1989) tested models of survival theory for the same. The model predicts the likelihood of a vehicle colliding at a time, assuming the vehicle survives until that time. This approach required specific data, which other researchers should have widely accepted.

Based on the above-mentioned literature studies, here in Table 1 shows the comparison table for statistical models used in accident prediction:

DOI: https://doi.org/10.52756/ijerr.2025.v47.007

Table 2. Comparison table for Machine Learning Approach.

Machine Learning Approach	Description	Application	References
Neuro-Fuzzy System	Integrates fuzzy logic and neural networks to infer meaning from ambiguous and inadequate data, aiding in recognizing hot traffic areas.	Identifying traffic hotspots by leveraging the unique abilities of Neuro-Fuzzy systems.	Huang and Ren (1999); Lee et al. (2003); Pappis and Mamdani (1977); Sayed et al. (1995)
Principal Component and Combination Grid Clustering	Utilizes principal component and combination grid clustering to locate high-density accident Black Spots on rural roadways, particularly in rapidly developing regions like rural China.	Identifying high-density accident Black Spots on rural roads with uncertain accident sites and complicated networks.	Shen et al. (2019)
Support Vector Machine (SVM) Strategy	Optimizes the learning of accident Black Spots using a support vector machine strategy, maximizing the classification interval to train a complex model for more precise detection of Black Spots.	Improving the precision of Black Spot detection through the optimization of SVM-based learning in accident prediction.	Fan et al. (2019)
Deep Neural Network (DNN)	Introduces a deep neural network-based approach for Black Spot identification, leveraging multi-source data on traffic accidents to build a feature-based detection system.	Locating Accident Black Spots using a deep neural network trained with relevant data category information.	Alagarsamy et al. (2021)
Adaptive Neuro-Fuzzy Inference System	Applies an adaptive Neuro-Fuzzy inference method to anticipate and pinpoint traffic hotspots on rural routes, combining a backpropagation algorithm with a least squares approach to train a fuzzy inference system.	Anticipating and pinpointing traffic hotspots on rural routes by integrating fuzzy logic and neural networks in an adaptive inference framework.	Hosseinlou and Sohrabi (2009)
Fuzzy Multicriteria Decision Making (FMCDM) and Fuzzy Rule-Based System (FRBS)	Utilizes FMCDM and FRBS techniques to calculate the Accidental Black location Severity Index (ABSSI) and Accidental Black Spot Severity Level (ABSSL) for categorizing and ranking accident Black Spots based on severity.	Categorizing and ranking accident Black Spots based on severity indices derived from FMCDM and FRBS techniques.	Bobade and Lad (2021)
Random Forest and Gaussian Distribution Methods	Employed to identify high-risk accident zones by providing alert information to users, potentially preventing accidents and saving lives.	Alerting users about high-risk accident zones on maps using a combination of random forest and Gaussian distribution methods.	Alagarsamy et al. (2021)

Machine Learning Approaches

On the other hand, new mathematical research, like the Neuro-Fuzzy system, shows that such tools may infer meaning from ambiguous and inadequate data because of their unique abilities. This feature of the Neuro-Fuzzy system can help with various transportation issues, such as recognizing hot traffic areas (Huang and Ren, 1999; Lee et al., 2003; Pappis and Mamdani, 1977; Sayed et al., 1995).

Shen et al. (2019) proposed employing principal component and combination grid clustering to locate high-density accident Black Spots on rural roadways. This method is primarily intended for nations like China, where rural roads are developing quickly. Gridding-based clustering simplifies the analysis of rural roads with

uncertain accident sites and complicated networks by condensing an object space to a limited road segment.

Based on the structured association properties of big data on urban traffic accidents, Fan et al. (2019) proposed a support vector machine strategy that optimizes the learning of accident Black Spots in the study domain and maximizes the classification interval used to train the complex model. The detection of Black Spots is now more precise. In addition, a deep neural network-based black spot identification method is introduced to deal with the expanding multi-source data on traffic accidents. To test the model's ability to locate accident Black Spots, a deep neural network is built with relevant data category information. Employing a deep neural network-based feature-based Black Spot detection system is advised.

Utilizing a comprehensive and integrated decision-making framework is the foundation for identifying accident factors. As a result, using conventional mathematical and statistical methods in this area may provide discrete problems. New research approaches that can decipher tangled and unclear data are helpful. Therefore, an adaptive Neuro-Fuzzy inference method finds applicability in anticipating and pinpointing traffic hotspots on rural routes. Using a hybrid optimization strategy created by combining a backpropagation algorithm with a least squares approach, a fuzzy inference system of the Sugeno type is trained in this procedure (Hosseinlou and Sohrabi, 2009).

This study suggests that fuzzy logic is the best approach for detecting accidental Black Spots. We may calculate the Accidental Black location Severity Index (ABSSI) and Accidental Black Spot Severity Level (ABSSL) of that specific location using Fuzzy Multicriteria Decision Making (FMCDM) and Fuzzy Rule-Based System (FRBS) techniques, respectively. The Accidental Black Spots are further categorized and ranked based on their severity index. Based on the severity level, it can be classified as highly severe, severe, just severe, or not severe (Bobade and Lad, 2021).

The areas most likely to be the scene of accidents are chosen and designated as black spots. Users receive alert information as a Black Spot on the map because they will likely use it while travelling to a new place. The risky or often accident-prone zone is mentioned for consumers travelling to a new site. Many people travelling in high-risk locations are saved from death by the proposed random forest approach and the Gaussian distribution method (Alagarsamy et al., 2021).

Based on the aforementioned research reviews, Table 2 shows a comparative table summarising the various Machine techniques for finding black spots and predicting traffic accidents.

Discussion

It is evident from earlier studies that several investigations have extensively used multiple linear regression models for accident prediction. This system might be enhanced in various ways to predict crash rates. The crash rates do not have a normal distribution, but this model assumes that the dependent variables do. Additionally, it falsely assumed no connection between a mistake and an independent variable. Additionally, this model generates negative values, which are impossible in the case of an accident.

Furthermore, accidents do not occur in a specific location. It predicts zero as the casualty rate, which contradicts the fact that a zero number indicates that the location is recognized as a safe location (Kim et al., 2005). Multiple linear regression models are typically built on the assumption of normally distributed errors. A numerical variable in this method defines the dependent variable.

Logistic Regression is a form in which the dependent is a categorical variable. It has been the most often used method for creating accident severity prediction models. Poisson models deal with discrete data and provide the best characteristics for representing automobile crashes as random numbers. These models, however, could provide inaccurate coefficients and standard errors if the data are very variable (Rahman, 2012). The model must be updated because most factors that explain the frequency of accidents are data categories. When the requirement of mean equal to variance becomes flexible, it was found that negative binomial regression models provided more desired properties than Poisson models for characterizing the association between accident occurrence and road features.

Simple, precise functions cannot gauge the seriousness of the problem, given that traffic accidents happen in real-time. Fuzzy-based systems create intricate non-linear correlations between parameters, which makes them ideal for predicting automobile collisions. However, a system that only used fuzzy logic would not be user-specific (Venkatesh et al., 2005). Neural networks are the best option because of their intelligent learning capabilities (Sarkar et al., 2012). The literature study indicates that hybrid approaches, such as Neuro-Fuzzy or Fuzzy-Neural methods, were preferred by researchers to conduct their investigations. Fuzzy techniques for clustering and neural network approaches are utilized to reveal the unique link inside each cluster. Both cooperative and concurrent NF models are not responsible due to the incorporation of ANN (black box notion). A hybrid NF model, on the other hand, is comprehensible and capable of learning under supervision (Abraham, 2001). As a result, many well-known hybrid NF models have been created, including Generalized Approximate Reasoning based Intelligence Control (GAIRC) and ANFIS. It usually takes work to conceptually compare the different neuro-fuzzy models and evaluate their effectiveness due to the need for a standardized framework (Abraham, 2001).

Neuro-fuzzy systems are generally intelligent. Quick learning, online flexibility, and minimal computational cost are crucial criteria. Neuro-fuzzy systems must also collect and prepare training data to be successful (Abraham, 2001). All Neuro-Fuzzy models employ gradient descent methods to discover the membership

DOI: https://doi.org/10.52756/ijerr.2025.v47.007

Table 3. Comparison table of strengths and weaknesses of various Statistical and Machine Learning techniques.

Model	Strengths	Weaknesses
Simple models based on mean and variance	Easy to implement	Ignores risk factors
Multiple Linear Regression models	Can incorporate speed and traffic volume	Assumes normal distribution (inapplicable for accident data)
Poisson distribution approach	Suitable for randomly occurring events	Limited to low-likelihood events
Poisson-gamma generalized linear model	Accounts for overdispersion	Complex model requires expertise
Poisson log-linear model	Provides insights into fluctuations	Ignores extra-Poisson variation
Negative binomial regression models	Addresses extra-Poisson variation	Requires specific data, not widely applicable
Empirical Bayesian approaches (EB)	Improves detection ability	Limited research on applicability
Neuro-Fuzzy System	Can handle ambiguous and incomplete data	Complex model requires expertise
Principal Component and Grid Clustering	Simplifies analysis of rural roads	Limited to specific contexts (e.g., rapidly developing rural areas)
Support Vector Machine (SVM)	Improves accuracy of black spot classification	Requires large datasets for training
Deep Neural Network (DNN)	Handles complex, multi-source data	Requires a lot of data and computational power
Fuzzy Logic with Multi-Criteria Decision Making (MCDM)	Provides severity levels for black spots	Subjective in defining fuzzy rules
Random Forest	Effective in identifying high-risk locations	Black box model, interpretability can be difficult
Gaussian Distribution Method	Can be used for risk assessment	Assumes normality of data (may not be applicable for accident data)
Neuro-Fuzzy System	Can handle ambiguous and incomplete data	Complex model requires expertise

function parameters. Examining more efficient neural network learning algorithms for accelerated learning and convergence will be beneficial. A neuro-fuzzy approach will deliver the desired result when the driver characteristics are used as a predictor variable in the accident prediction model (Sarkar et al., 2016; Sarkar et al., 2012).

Based on the aforementioned research reviews, a comparative Table 3 summarises the strengths and weaknesses of various Statistical and Machine Learning techniques for finding black spots and predicting traffic accidents

Conclusion

Traffic accident prediction is essential for the urban traffic accident prediction model and for better traffic safety management. Several methods have been applied to predict accidents on urban roadways. While some use statistical methods, others more recently use soft computing approaches. Fuzzy-based algorithms are ideal for predicting vehicle collisions because they can readily construct complicated non-linear correlations between variables. However, a fully fuzzy-based strategy would provide the user-defined insight required. Neural Networks are ideal for this requirement through their intelligent learning capabilities.

Conflict of interest

The authors declare no conflict of interest.

References

Abraham, A. (2001). Neuro fuzzy systems: State-of-the-art modeling techniques. In J. Mira & A. Prieto (Eds.), *Connectionist Models of neurons, Learning processes and Artificial Intelligence, 2084*, 269–276. Springer. https://doi.org/10.1007/3-540-45720-8_30

Abdel-Aty, M. A., & Radwan, A. E. (2000). Modelling traffic accident occurrence and involvement. *Accident Analysis & Prevention, 32*(5), 633–642. https://doi.org/10.1016/S0001-4575(99)00094-9

Alagarsamy, S., Malathi, M., Manonmani, M., Sanathani, T., & Senthil, K. A. (2021). Prediction of road accidents using machine learning technique. In *2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, pp. 1746–1750.

https://doi.org/10.1109/ICECA52323.2021.9675852

Alhosani, K., & Alhashmi, S.M. (2024). Opportunities, challenges, and benefits of AI innovation in government services: a review. *Discov Artif Intell*. *4*, 18. https://doi.org/10.1007/s44163-024-00111-w

Blower, D., Campbell, K., & Green, P. (1993). Accident rates for heavy truck-tractors in Michigan. *Accident Analysis & Prevention, 25*(3), 307–321. http://hdl.handle.net/2027.42/30768

Bobade, S. U., & Lad, R. K. (2021). Accidental black spot analysis by fuzzy approach: A review. In *Proceedings of the International Conference on IoT Based Control Networks & Intelligent Systems - ICICNIS 2021*. https://dx.doi.org/10.2139/ssrn.3884055

Brown, B. C., & Farley, M. F. (1992). Identification of dangerous highway locations: Results of community health department study in Quebec. *Transportation Research Record, 1376*, 78–85. https://onlinepubs.trb.org/Onlinepubs/trr/1992/1376/1376-011.pdf

Caliendo, C., Guida, M., & Parisi, A. (2007). A crash-prediction model for multilane roads. *Accident Analysis & Prevention, 39*(4), 657–670. https://doi.org/10.1016/j.aap.2006.10.012

Ceder, A., & Livneh, M. (1982). Relationships between road accidents and hourly traffic flow. *Accident Analysis & Prevention, 14*(1), 19–34. https://doi.org/10.1016/0001-4575(82)90004-5

Cheng, L., Geedipally, S. R., & Lord, D. (2013). The Poisson–Weibull generalized linear model for analyzing motor vehicle crash data. *Safety Science, 54*, 38–42. https://doi.org/10.1016/j.ssci.2012.11.002

Fan, Z., Liu, C., Cai, D., & Yue, S. (2019). Research on black spot identification of safety in urban traffic accidents based on machine learning method. *Safety Science, 118*, 607–616. https://doi.org/10.1016/j.ssci.2019.05.039

Foldvary, L. (1979). Road accident involvement per miles travelled—V. *Accident Analysis & Prevention, 11*(2), 75–99. https://doi.org/10.1016/0001-4575(79)90017-4

Geurts, K., & Wets, G. (2003). *Black spot analysis methods: Literature review* (p. 30). Universitaire Campus GEBOUW D, Diepenbeek. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=a3e4397b86458573e69ed50381f21cb31a584018

Hauer, E. (2001). Overdispersion in modelling accidents on road sections and in empirical Bayes estimation. *Accident Analysis & Prevention, 33*(6), 799–808. https://doi.org/10.1016/S0001-4575(00)00094-4

Hauer, E. (1997). Observational before–after studies in road safety: Estimating the effect of highway and traffic engineering measures on road safety. Pergamon. https://trid.trb.org/View/472973

Hauer, E., Ng, J. C. N., & Lovell, J. (1988). Estimation of safety at signalized intersections. *Transportation Research Record, 1185*, 48–61. https://onlinepubs.trb.org/Onlinepubs/trr/1988/1185/1185-005.pdf

Hauer, E., & Persaud, B. N. (1987). How to estimate the safety of rail-highway grade crossings and the effects of warning devices. *Transportation Research Record, 1114*, 131–140. http://onlinepubs.trb.org/onlinepubs/trr/1987/1114/1114-015.pdf

Heydecker, B. G., & Wu, J. (2001). Identification of sites for accident remedial work by Bayesian statistical methods: An example of uncertain inference. *Advances in Engineering Software, 32*(10–11), 859–869. https://doi.org/10.1016/S0965-9978(01)00037-0

Higle, J. L., & Witkowski, J. M. (1988). Bayesian identification of hazardous locations. *Transportation Research Record, 1185*, 24–36. http://onlinepubs.trb.org/onlinepubs/trr/1988/1185/1185-003.pdf

Hosseinlou, M. H., & Sohrabi, M. (2009). Predicting and identifying traffic hot spots applying neuro-fuzzy systems in intercity roads. *International Journal of Environmental Science & Technology, 6*(2), 309–314. https://doi.org/10.1007/BF03327634

Huang, H., Chin, H. C., & Md. Mazharul, H. (2008). Severity of driver injury and vehicle damage in traffic crashes at intersections: A Bayesian hierarchical analysis. *Accident Analysis & Prevention, 40*(1), 45–54. https://doi.org/10.1016/j.aap.2007.04.002

Huang, S., & Ren, W. (1999). Use of neural fuzzy networks with mixed genetic or gradient algorithm in automated vehicle control. *IEEE Transactions on Industrial Electronics, 46*(6), 1090–1102. https://doi.org/10.1109/41.807993

Jovanis, P., & Chang, H. (1989). Disaggregate model of highway accident occurrence using survival theory. *Accident Analysis & Prevention, 21*(5), 445–458. https://doi.org/10.1016/0001-4575(89)90005-5

Jovanis, P., & Delleur, J. (1983). Exposure-based analysis of motor vehicle accidents. *Transportation Research Record, 910*, 1–7. http://onlinepubs.trb.org/Onlinepubs/trr/1983/910/910-001.pdf

Karamanlis, I., Kokkalis, A., Profillidis, V., Botzoris, G., & Galanis, A. (2023). Identifying Road Accident Black Spots using Classical and Modern Approaches. *Wseas Transactions on Systems, 22*, 556–565. Portico. https://doi.org/10.37394/23202.2023.22.56

Kim, D. G., Washington, S., & Oh, J. (2006). Modeling crash types: New insights into the effects of covariates on crashes at rural intersections.

Journal of Transportation Engineering, 132(4), 282–292. https://doi.org/10.1061/(ASCE)0733-947X(2006)132:4(282)

Kim, S. H., Chung, S. B., Song, K. H., & Chon, K. S. (2005). Development of an accident prediction model using GLIM (Generalized Log-linear Model) and EB method: A case of Seoul. *Journal of the Eastern Asia Society for Transportation Studies, 6*, 3669–3682. https://doi.org/10.11175/easts.6.3669

Korhonen, J. M. (2018). Overcoming Scarcities Through Innovation: What Do Technologists Do When Faced with Constraints? *Ecological Economics, 145*, 115-125. https://doi.org/10.1016/j.ecolecon.2017.08.023

Kristen, L. S. B., & Mark, R. V. (2002). Improving the identification, analysis and correction of high-crash locations. *ITE Journal, 72*(1), 38–43.

Kulmala, R. (1995). *Safety at rural three- and four-arm junctions*. Technical Research Centre of Finland (VTT), Espoo, Finland.

Lee, C., Hellinga, B., & Saccomanno, F. (2003). Proactive freeway crash prevention using real-time traffic control. *Canadian Journal of Civil Engineering, 30*(6), 1034–1041. https://doi.org/10.1139/l03-040

Li, W., Carriquiry, A., Pawlovich, M., & Welch, T. (2008). The choice of statistical models in road safety countermeasure effectiveness studies in Iowa. *Accident Analysis & Prevention, 40*(4), 1531–1542. https://doi.org/10.1016/j.aap.2008.03.015

Maher, M. J., & Summersgill, I. (1996). A comprehensive methodology for the fitting of predictive accident models. *Accident Analysis and Prevention, 28*(3), 281–296. https://doi.org/10.1016/0001-4575(95)00059-3

Maycock, G., & Hall, R. D. (1984). *Accidents at 4-arm Roundabouts*. *Transport Research Laboratory Report*, 1120.

Meel, B. L., & Safr, M. J. (2008). Fatal road traffic accidents in the Mthatha area of South Africa. *South African Medical Journal*, 98(9), 716–719. http://www.samj.org.za/index.php/samj/article/view/146/2071

Miaou, S. P. (1994). The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. *Accident Analysis and Prevention, 26*(4), 471–482. https://doi.org/10.1016/0001-4575(94)90038-8

Mitra, S., & Washington, S. (2007). On the nature of over-dispersion in motor vehicle crash prediction models. *Accident Analysis and Prevention, 39*(3), 459–468. https://doi.org/10.1016/j.aap.2006.08.002

Nassar, S. A. (1996). *Integrated road accident risk model* (Doctoral dissertation). Waterloo University, Ontario, Canada.

Norden, M., Orlansky, J., & Jacobs, H. (1956). Application of statistical quality-control techniques to analysis of highway accident data. *Highway Research Board Bulletin, 117*, 17–31. http://onlinepubs.trb.org/Onlinepubs/hrbbulletin/117/117-002.pdf

Oppe, S. (1979). The use of multiplicative models for analysis of road safety data. *Accident Analysis and Prevention, 11*(2), 101–115. https://doi.org/10.1016/0001-4575(79)90018-6

Pappis, C., & Mamdani, E. (1977). A fuzzy controller for a traffic junction. *IEEE Transactions on Systems, Man, and Cybernetics, 7*(17), 707–717.

Persaud, B. N., Lyon, C., & Nguygen, T. (1999). Empirical Bayes procedure for ranking sites for safety investigation by potential safety improvements. *Transportation Research Record, 1665*, 7–12. https://doi.org/10.3141/1665-02

Persaud, B. N. (1990). Black spot identification and treatment evaluation. The Research and Development Branch, Ontario Ministry of Transportation. https://trid.trb.org/View/356660

Petridou, E., & Moustaki, M. (2000). Human factors in the causation of road traffic crashes. *European Journal of Epidemiology, 16*(9), 819–826. https://doi.org/10.1023/A:1007649804201

Rashid, A. B., & Kausik, M. A. K. (2024). AI revolutionizing industries worldwide: A comprehensive overview of its diverse applications. *Hybrid Advances, 7*, 100277. https://doi.org/10.1016/j.hybadv.2024.100277

Quddus, M. A. (2008). Time series count data models: An empirical application to traffic accidents. *Accident Analysis and Prevention*, 40(5), 1732–1741. https://doi.org/10.1016/j.aap.2008.06.011

Rahman, S. (2012). Development of an accident prediction model for intersections of Dhaka City, Bangladesh. *International Journal of Computer Applications, 47*(16), 10–16. https://doi.org/10.5120/7271-0389

Saccomanno, F., & Buyco, C. (1988). Generalized loglinear models of truck accident rates. *Transportation Research Record, 1172*, 23–31. https://onlinepubs.trb.org/Onlinepubs/trr/1988/1172/1172-003.pdf

Sarkar, A., Sahoo, G., & Sahoo, U. C. (2016). Feature selection in accident data: An analysis of its application in classification algorithms. *International Journal of Data Analysis Techniques and Strategies, 8*(2), 108–121. https://doi.org/10.1504/IJDATS.2016.077484

Sarkar, A., Sahoo, U. C., & Sahoo, G. (2012). Accident prediction models for urban roads. *International Journal of Vehicle Safety, 6*(2), 149–161. https://doi.org/10.1504/IJVS.2012.049020

Sayed, T., Abdelwahab, W., & Navin, F. (1995). Identifying accident-prone locations using fuzzy pattern recognition. *Journal of Transportation Engineering*, *121*(4), 352–358. https://doi.org/10.1061/(ASCE)0733-947X(1995)121:4(352)

Shankar, V., Mannering, F. L., & Barfield, W. (1995). Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. *Accident Analysis and Prevention*, *27*(3), 371–389. https://doi.org/10.1016/0001-4575(94)00078-Z

Shen, L., Lu, J., Long, M., & Chen, T. (2019). Identification of accident blackspots on rural roads using grid clustering and principal component clustering. *Mathematical Problems in Engineering*, *2019*(4), 1–12. https://doi.org/10.1155/2019/2151284

Summersgill, I., & Layfield, R. E. (1996). Non-junction accidents on urban single-carriageway roads (TRL Report No. 183). Transport Research Laboratory.

Tamburri, T. N., & Smith, R. N. (1970). The safety index: Method of evaluating and rating safety benefits. In *Highway Research Record* (Vol. 332, pp. 28–43). Highway Research Board, National Research Council. http://onlinepubs.trb.org/Onlinepubs/hrr/1970/332/332-003.pdf

Venkatesh, K., Ramesh, A., Alagusundaram, M., & Sahana, J. V. (2005). NEFCOP: A neuro-fuzzy vehicle collision prediction system. In *Proceedings of the International Conference on Computational Intelligence for Modelling, Control and Automation*, Vienna. https://doi.org/10.1109/CIMCA.2005.1631406

Vistisen, D. (2002). *Models and methods for hot spot safety work* (Doctoral dissertation). Technical University of Denmark. http://www2.imm.dtu.dk/pubdb/edoc/imm889.pdf

Vonderohe, A., & Chou, C. (1997). *A generic data model for linear referencing systems*. National Cooperative Highway Research Program. Transportation Research Board. https://onlinepubs.trb.org/onlinepubs/nchrp/nchrp_rrd_218.pdf

Wan, Y., He, W., & Zhou, J. (2021). Urban Road Accident Black Spot Identification and Classification Approach: A Novel Grey Verhuls–Empirical Bayesian Combination Method. *Sustainability*, *13*, 11198. https://doi.org/10.3390/su132011198

How to cite this Article:

Amrita Sarkar and Vandana Bhattacharjee (2025). A Comparative Study of Accident Black Spot Identification: A Crucial Analysis Using Machine Learning and Statistical Approach. *International Journal of Experimental Research and Review*, *47*, 84-92.

DOI : https://doi.org/10.52756/ijerr.2025.v47.007

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.